DOCUMENT RESUME

ED 455 791 IR 020 747

AUTHOR Schnackenberg, Heidi L.; Chin, Kevin; Luppicini, Rocci J.
TITLE Heuristic and Formative Evaluation: A Case Study Illustrator of a New Technique.
PUB DATE 2000-10-00
NOTE 9p.; In: Annual Proceedings of Selected Research and Development Papers Presented at the National Convention of the Association for Educational Communications and Technology (23rd, Denver, CO, October 25-28, 2000). Volumes 1-2; see IR 020 712.
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Case Studies; Computer Assisted Instruction; *Evaluation Methods; *Formative Evaluation; *Heuristics; Instructional Design; *Instructional Effectiveness; *Instructional Material Evaluation; Screen Design (Computers); World Wide Web
IDENTIFIERS *Web Sites

ABSTRACT

As the value of media-delivered information is increasingly emphasized as a powerful instructional tool, user knowledge gains are being more frequently investigated. Formative evaluation techniques that assess learning are often employed. While it is important to appraise this aspect of an instructional program or Web site, it is also necessary to investigate the functionality of the electronic characteristics of the same program. Heuristic evaluation techniques are rapidly becoming the evaluation method of choice to assess this aspect of instructional programs. The two areas-functionality and learning-are rarely both assessed within the same evaluation due to time, money, and methodological constraints. A new type of evaluation needs to evolve to keep current with new design and development strategies. The purpose of this study is to further investigate the use of heuristic evaluation, in conjunction with formative evaluation, as a methodology for assessing instructional Web sites. Through this type of mixed methodology, not only can attitude, user interface, and navigation issues be explored, learning and instructional strategies can be investigated as well. An evaluation of this nature, illustrating the proposed methodology, is described in the case study discussed in this paper. (Contains 19 references.) (Author/AEF)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 455 791

Heuristic and Formative Evaluation: A Case Study Illustrator of a New Technique

Heidi L. Schnackenberg
Kevin Chin
Rocci J. Luppicini
Concordia University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

M. Simonson

Abstract

In recent decades, instructional computer programs and websites have become increasingly prevalent. Programmers, academics, and free-lance computer professionals have all begun to create what they tout to be instructionally sound products that will help people learn. Many of these products however, are rarely evaluated on any level for various facets of their design. When they are evaluated, it is usually either to gather information about the navigability or functioning of the hardware/software, or even less frequently, to assess how well people have acquired the skills and knowledge taught in the program.

As the value of media-delivered information is increasingly emphasized as a powerful instructional tool, user knowledge gains are becoming more frequently investigated. Formative evaluation techniques that assess learning are often employed. While it is important to appraise this aspect of an instructional program or website, it is also necessary to investigate the functionality of the electronic characteristics of the same program. Heuristic evaluation techniques are rapidly becoming the evaluation method of choice to assess this aspect of instructional programs. Unfortunately, the two areas - functionality and learning - are rarely both assessed within the same evaluation due to time, money, and methodological constraints.

While the growing popularity of media-based instructional programs has advanced instructional design and development techniques, equally efficient and effective evaluation methodologies to correspond to this new manner of instructional design have fallen behind. Current evaluation techniques for electronically delivered instruction are either poor in their methodology or incomplete in their design. Often, all pertinent aspects of instructional programs are not assessed and the learner is left with only a partially-sound form of instruction. A new type of evaluation therefore needs to evolve to keep current with new design and development strategies. A hybrid of heuristic and formative evaluation is proposed.

"Evaluation is a discipline inquiry to gather facts and other evidence that allow an evaluator to make assertions about the quality, effectiveness or value of a program, a set of materials, or some other object of the evaluation in order to support decision making" (Cummings, 1998). As such, evaluation techniques and methodologies exist in many different fields, and are performed on a wide variety of materials, programs, and products. While evaluation in its purest concept adheres to Cummings (1998) definition, particular types of evaluation are performed with specific goals in mind, using unique methodologies. Heuristic and formative evaluations are among the two more prevalent methods used in the field of educational technology at present.

Heuristic Evaluation

Heuristic evaluation is a type of usability testing. Usability testing has its roots in classical experimental methodology (Rubin, 1994) and has experienced popular application in the field of engineering. It is a systematic way of evaluating the functionality of a product (usually electronic) by observing users and recording information about areas of difficulty and ease within a program (Dumas & Redish, 1993). Dumas and Redish (1993) describe five characteristics of every usability test as follows: 1) the primary goal is to improve product usability, 2) individual usability tests also have unique goals that are determined based on specific needs, 3) testing evaluators are actual users, 4) testing evaluators perform authentic tasks, 5) problem areas are revealed through data analysis and modifications are recommended.

Commonly, usability testing is not implemented in its purest form (Nielsen 1993; Whiteside, Bennett, & Holtzblatt, 1988). Costs for full-scale usability tests are perceived to be very prohibitive (Nielsen, 1994), and their methodologies very complex (Belotti, 1988). A simplified method of usability was therefore developed by Nielsen in 1989 called "discount usability engineering" (Nielsen 1989b, 1990a, 1993). One of the most popular types of discount usability engineering is termed "heuristic evaluation."

IR020747

 BEST COPY AVAILABLE

Heuristic evaluation (Nielsen 1994) engages a small set of evaluators, usually three to five, to examine an interface and assess its adherence to a prespecified set of usability criteria, or "heuristics". Each evaluator progresses through a program individually and either records their findings in writing or verbalizes them to an observer who is present during each session. If an observer is present, he/she is also allowed to assist the evaluator in navigating the site if necessary or to answer other questions that may arise. A debriefing session may also occur at the end of each evaluation session to gather more information. After all the evaluators have completed their sessions, the data is then analyzed and items needing to be revised are identified.

Formative Evaluation

In many respects, heuristic evaluation is not unlike formative evaluation as classically defined in the educational technology literature. Coined in 1967 by Michael Scriven, the term formative evaluation is viewed as a means of identifying areas of modification in the development of educational materials through the collection and analyzation of data from the target population. This is different from "summative evaluation" which occurs after development in order to determine effectiveness (Smith & Ragan, 2000).

Dick and Carey (1996) propose three phases to formatively evaluating instructional materials. These are the one-to-one or clinical evaluation, the small-group evaluation, and the field trial. The one-to-one evaluation stage occurs individually with one to three learners who are representative of the target population. Ideally the three consist of one high-ability, one medium-ability and one low-ability learner. The one-to-one evaluation is utilized in order to identify any factual errors in the instruction and to obtain initial reactions and indications of performance improvement. Questionnaires regarding learner attitudes are generally used as the main data collection instrument.

Once the instructional materials have been revised according to the information gathered from the one-to-one evaluation, a small group evaluation should be performed with approximately eight to twenty learners. Again, these learners should be representative of the target population as much as possible. Learners should be selected at random so that your results can be generalized to the entire population. Two primary purposes for the small group evaluation are to 1) determine the effectiveness of changes made following the one-to-one evaluation, and 2) identify any remaining learning problems that learners may have. In this phase, learner performance scores on pretests and posttests are typically used to evaluate instructional effectiveness. Attitudes toward the instruction are evaluated through questionnaires or follow-up interviews.

The field trial is the final formative evaluation phase that Dick and Carey (1996) discuss. It involves the participation of a randomly sampled group of about thirty individuals who are representative of the target population. It is used to determine the effectiveness of the changes resulting from the small group evaluation and whether the instruction can be used in the context for which it is intended. Much like a dress rehearsal, it provides the last chance to identify and remove any remaining errors or problems. There are many similarities between the field trial and the small group evaluation. The main difference between the two evaluation processes is in the actual authenticity of the materials, learners, procedures, instructors and setting. The field trial should mirror the intended instructional experience as much as is possible.

Several similarities exist between heuristic and formative evaluation. Some experts consider formative evaluation to be the underlying blueprint for heuristic evaluation (Hix & Hartson, 1994). The two methodologies also appear to have similar goals. They both use data collected from a target population in order to make recommendations regarding modifications to a specific product or material in the design and development phases of creation. Both employ the use of surveys, observations, interviews, and various other data collection instruments and techniques. However, despite apparent similarities, there do exist fundamental differences between the two. Formative evaluation primarily focuses instructional and learning strategies, as evidenced by the use of pretests and posttests. Heuristic evaluation concentrates more on the usefulness of a product, i.e. the user interface, navigation issues, etc.

In 1987, Patterson and Bloch called for formative evaluation to be conducted during the development of computer-assisted instruction (CAI). In their article, they propose investigating learning gains, user attitudes, interface and navigation issues utilizing Dick and Carey's (1996) three phases of formative evaluation as a structure. These areas of investigation within educational media products still remain of utmost importance today. However, with the advent of the Internet and the World Wide Web (WWW), as well as the ever-increasing rapidity with which electronic educational products are produced, the implementation of the methodology Patterson and Bloch (1987) propose is impractical in the currently fast-paced realm computerized instruction.

A methodology that would perhaps serve the needs of efficiency, practicability, and still investigate learning, attitudes, interface and navigation therefore needs to be created. The present work attempts to address such a problem with the development and testing of a new method. It consists of a combination of heuristic and formative evaluation techniques. In 1997, Corry, Frick, and Hansen incorporated usability testing techniques into

the development of university informational website. They used many facets of heuristic evaluation as the structure of their investigation. According to their results, this was a very efficient method for exploring interface, navigation, content and some attitudinal issues. Learning, however, could not be measured as the site was informational and not instructional.

The purpose of the present study was to further investigate the use of heuristic evaluation, in conjunction with formative evaluation, as a methodology for assessing instructional websites. Through this type of mixed methodology, not only can attitudes, user interface, and navigation issues be explored, learning and instructional strategies can be investigated as well. An evaluation of this nature, illustrating the proposed methodology, is described in the case study that follows.

Method

Participants

Participants in the study were five students at an urban university in or prior to their first semester in a graduate programme in education. All participants will soon be formally enrolled in a required Human Performance Technology course and participated in the study because the instructional program used in it covers content regularly taught in the course. Although the site content was required course material, the evaluation sessions were several hours long and it was decided that evaluators would be paid a small stipend for their time.

Materials

The instructional materials were designed to teach various Human Performance concepts, including needs analysis, instructional design, and formative and summative evaluation. Actual course content and activities were adapted into an interactive instructional website for the study. The site was created initially with the intention of supplementing an on-site course. The long term goal of the website is to have it serve as a stand-alone, distance-delivered web course.

The site contained a total of 10 instructional modules, a homepage (including the course syllabus and navigation information), and a sitemap. Eight of the ten modules contained multiple-choice and/or constructed-response practice-with-feedback activities. The remaining two modules were solely informational. For the current purposes, the site is supplemental to the on-site Human Performance Technology course. In future, it will be used as the basis for a distance version of the same course. The current URL for the site is http://doe.concordia.ca/etec512_712/index.html

The website contained three overall course objectives and 23 learning objectives. 16 of these objectives were short-term and to be completed within the site, while seven of the objectives were subobjectives of the long-term course objectives. Each of the objectives was taught through a number of screens which presented instruction, five practice-with-feedback items (for the short-term objectives), summaries, and reviews. Seven objectives required selected responses in a multiple-choice format and nine required constructed responses. Practice items consisted of multiple-choice questions with two-to-four response choices for the seven selected-response objectives. For the constructed-response items, participants typed their answer in a field and pressed a submit button. In the next field a sample answer appeared to which participants compared their answer. The site was unable to track each participant's progress by recording response choices due to the inability of the university server to support the necessary interface.

Learners could advance through the site by selecting links to modules from the menu or site map. Once in a module, they could choose to view any screen by selecting a link from a table of contents, or they could simply advance in a linear fashion by clicking "previous" or "next" buttons that appeared on each screen. As the site was programmed to work on either a PC or a Macintosh platform, learners were also able to progress through the site using the navigation options within the web-browser they were utilizing.

Procedures

Prior to engaging in the evaluation of the instructional website, it was ascertained that participants were graduate students in the department of education, who were enrolled in a degree programme. The participants also must not have taken the Human Performance Technology course prior to evaluating the website. These were the only prerequisites necessary for learners to be able to evaluate the website. Permission was given by the participants to video-tape each evaluation session in case the tapes of the sessions had to be reviewed for data gathering purposes.

During each session, each participant was then stationed at a computer terminal and given access to the website. He or she was presented with a list of heuristics and instructed to record errors and respond to the various heuristics while progressing through the site. Participants were also asked to verbalize their thoughts and questions

while utilizing the program (Smith & Wedman, 1988). A researcher sat beside each participant, recording observations as he/she advanced through the site. The researcher answered any questions that arose regarding navigation or content, although he redirected the participants to try to answer content-related questions by referring to the site. Also, the participants were asked questions in response to their comments in order to probe more deeply into the nature of participant statements. Full learner control was given to the participants due to the fact that they could progress through the site at any pace they chose, viewing any screens at any point, with the only restriction being that they must complete all the practice activities that correspond to the short-term objectives.

As the stated previously, the university server would not support any interface that would allow tracking or recording of participant progression through the site, so another method was devised to record student responses to the practice items. After each practice set, the researcher copied and pasted participant answers to the practice items into a separate word document. Correct or incorrect answers to the multiple-choice items were indicated, whilestatements from the participants as to the correctness of their answers in comparison to the sample answers for the constructed-response items were also recorded.

Overall, it took each participant approximately 11 hours to complete the evaluation of the website. As the evaluation was time consuming, participants completed the activity in several sessions which varied in length according to their individual preferences and schedules. Upon completion of the evaluation, each participant was interviewed in a short debriefing session to get any last impressions or thoughts he or she may want to communicate. A paper-and-pencil survey regarding attitudes toward the website, perceived future usefulness, and areas of modification was also administered at this time and participants returned the survey prior to leaving the session.

Criterion Measures

Learner achievement was measured through performance on the practice items. The 35 multiple-choice items were scored either one or zero, and the 45 constructed-response items were scored either two (completely correct answer), one (partially correct answer), or zero (no answer or incorrect answer) according to a scoring key developed by the experimenters. Thus, the maximum possible score on the practice items was 125.

The 32-item attitude questionnaire assessed participants' satisfaction with the material, their perceived future usefullness of the skills learned in the site, and suggestions for modifications. The attitude questionnaire, containing 12 three-choice Likert-type items, seven yes/no items, and 13 constructed-response items, was administered immediately after participants completed the instructional program. A sample item from the attitude questionnaire is below:

Circle how much you liked each of the activities listed below.

4. Receiving feedback that asked you to compare your answers to a sample answer.	I liked this a lot	I liked this OK	I did not like this

Responses to the list of heuristics were open-ended. There were five catagories in which to respond, including site content, site navigation, graphical appropriateness, readability, and communication venues. A sheet where general errors were recorded was included. Sample heuristics are presented below:

Please provide specific feedback (positive and/or negative) on the HPT website regarding the following:

- Site content (information, samples/examples, practice items, answers to practice items, links/resources)
- Site navigation (interface, platform conventions, i.e. buttons, etc., menubar, site map, navigation instructions)

Researcher observations to the think-aloud protocols were recorded per individual participants, per session. Observations were made in a variety of catagories. These catagories were quite similar to the heurstics and included site content, site navigation, graphical appropriateness, readability, and communication venues. Observations on learning and attitudes were also recorded.

Data Analysis

Calculation of simple mean scores for the practice items, individually and collectively, were tabulated to indicate achievement. Responses to the attitude survey were tallied and, for the open-ended items, categories of responses were created. Categories of reponses were also created to calculate data on the debriefing responses, heuristic lists and researcher observations.

Results

Achievement

Mean overall practice item scores per instructional objective are shown in Table 1. Within the "Interventions" module of the Web site, participants scored differently on the two module objectives. Although

participants scored relatively high when asked to classify performance solutions as instructional or non-instructional using a forced-choice format (84%), they scored significantly lower (46%) when prompted to offer possible solutions to given case scenarios using an open-ended question format. Another noteworthy result occurred within the "Practice and Feedback" module of the Web site. When prompted to identify appropriate practice activities for given objectives using a forced-choice format, a perfect mean overall score occurred (100%). Asked to write appropriate practice activities for given objectives, overall participant score lowered significantly (60%). The trend of participants scoring highly on closed-ended questions does not continue with the practice items score from the "Sequencing" module. Participants were presented with various course or workshop and asked to identify their pedagogical components. Participants answered all prompts incorrectly, resulting in an overall mean score of zero (0%). The overall percentage scores of participants ranged from 61.6% to 68.8%, with a mean overall percentage score of 65.3%.

Table 1
Mean Overall Practice Item Scores per Instructional Objective

Instructional Objectives	Mean Score (%)
Classify performance solutions as instructional or non-instructional.	84%
Offer possible solutions to given case scenarios.	46%
Develop sections of responses to given proposal scenarios.	58%
Given various data sets from a needs analysis, develop recommendations based on your conclusions from the data.	68%
Develop sample data collection items from given case scenarios.	68%
Identify well-written instructional objectives .	92%
Write instructional objectives.	66%
Identify appropriate assessment items for given instructional objectives.	88%
Identify well-written assessment items.	72%
Write appropriate assessment items for instructional objectives.	80%
Identify appropriate practice activities for given objectives.	100%
Write appropriate practice activities for given objectives.	60%
Identify the pedagogical components, given various courses or workshops.	0%
Determine whether the evaluations described in given scenarios are formative or summative.	92%
Design a methodology and state the instruments to be used for the type of evaluation indicated in given scenarios.	76%

Note: Total number of Participants = 5.

Attitudes

When asked to indicate liking of activities, participants responded positively to receiving feedback that asked to compare their answers to sample answers, with a majority of participants (80%) saying they liked the activity "a lot". When asked to indicate the importance of activities, all participants (100%) answered that both the reading material presented and relating the information in the Web site to future, practical applications was extremely important. A majority of participants (80%) answered that completing the practice exercises in the module were extremely important.

All participants (100%) did not believe the Web site was too hard for them to understand and complete. All participants (100%) also responded that they were able to successfully compare their answers to the sample answers and they learned important techniques that would be of value in the real world. A majority of participants (80%) found the practice exercises helpful, were able to successfully navigate the Web, relate the information presented to previous learning experiences and felt they would be able to apply what they learned in the Web site to a real world setting. When prompted for the most preferred topic of the Web site, participants mentioned the "Needs Assessment" module the most stating it was relevant, informative and easy to understand. When prompted for the least preferred topic, participants mentioned the "Objectives and Assessment" module the most, stating the practice items as the deciding factor.

Heuristic Commentaries

Participant commentaries per heuristic are shown in Table 2. When asked to remark about the content of the Web site, participants mentioned its unclear and/or confusing wording the most (31% of total heuristic comments made). Despite this, participants also stated that the Web site was clear and useful information was provided (25% of total heuristic comments made). Participants were satisfied with navigation issues (32% of total heuristic comments made). The graphics and reading level was judged as appropriate for the target audience (52% of total heuristic comments made). When prompted for errors within the Web site, the majority of statements (54% of total heuristic comments made) indicated grammatical errors in the content. The grammatical errors may also be a factor in participants stating that the content was confusing (19% of total heuristic comments made).

Table 2
Participant Commentaries per Heuristic

Heuristic	Number of Responses
Content	
Unclear or confusing wording.	(10)
Clear and useful information provided.	(8)
Examples with feedback very helpful.	(6)
Inconsistent presentation of content.	(4)
Insufficient information provided.	(4)
Navigation	
Satisfied with navigation issues.	(8)
Inconsistent presentation of content	(5)
Poor design issues.	(4)
Attractive site layout.	(2)
Non-working functions.	(1)
Graphics and Reading Level	
Appropriate reading level.	(11)
Text cut off on screen.	(4)
Insufficient graphics.	(2)
Confusing language used.	(2)
Poor design detracts from learning.	(2)
Communication Avenues	
Hyperlinks useful.	(5)
Sufficient for site.	(4)
Miscellaneous	(3)
Other	
Miscellaneous	(6)
Sufficient practice and feedback.	(3)
Informative content.	(1)
Poor design.	(1)
Poorly written content.	(1)
Errors	
Grammatical errors.	(60)
Confusing content.	(21)
Formatting errors.	(11)
Layout/design errors.	(8)
Miscellaneous	(5)
Non-working functions.	(2)

Observations

When prompted for comments related to navigation of the Web site, many related to the poor interface/layout design of the Web site. Participants were also vocal about the poor presentation of the content when

asked about the graphics and reading level. Many comments about the communication avenues within the Web site mentioned the its high-level access to information (within the site and links to external information sources.) The majority of comments relating to errors in the site related to inconsistent design features and confusing presentation of its content. When asked about the learning and instruction aspects of the Web site, most comments referred to the relevant practice and feedback examples.

Recommendations

Data from achievement on the practice items indicate that participants learned approximately 65% of the content included in the website. This would indicate that while the website appears to be an effective supplement to a course in human performance technology, it cannot at the present time be used as a stand-alone distance delivered web course. An investigation of scores per objective reveals that while participants generally attained successful scores on the selected-response practice items, scores were marginally lower on the constructed-response questions. Participants scored below 60% on objectives where they had to offer solutions to case scenarios, develop proposal sections, and write practice activities. The lowest achievement score however occurred where participants had to identify pedagogical components of an instructional design package. Attitude responses indicated that the directions for this section were unclear and that may account for the 0% success rate with these items. Conversely, attitude data also revealed that participants thought they were quite successful in comparing their answers to given sample answers. This may mean that they thought they scored well on these items but actually did not, or that they were successful in estimating that the answers they created were incorrect compared to the sample. What is clear from this data is that several of the constructed-response practice items need to be modified and that the directions for the practice items on pedagogical components need to be clarified.

Participant attitude data indicated that overall, participants like the site and thought that many of the activities in which they engaged were important. As well, participants responded that the material they learned would be useful in a "real world setting" and the skills they learned could be transferred to situations outside of the web environment. This data is revealing regarding future student potential to engage in learning the information in the website. Distance-delivered web-courses classically suffer high attrition rates. One factor that this is attributed to is motivation level. If students appear to like the website described in this study, as well as see value and transferability in the material, then it is likely that at least motivationally, the site would be a success if it is used as a stand-alone course. As a course supplement, attitude data indicates that it is more than engaging motivationally.

Heuristic response data was successful in indicating various errors within the website. The specific errors were not indicated here as they were not deemed of interest or import to anyone beyond the developers. What is of interest however is the fact that the technique of using five evaluators was successful in finding and documenting various website errors. Further, more general heuristic data revealed that while the reading level within the site was appropriate, the way certain concepts were presented was confusing and needs to be clarified for future use. Navigation issues did not appear to be of great concern as participants responded positively to the methods and venues provided for navigation within the site.

Finally, participant think-aloud responses further iterated the need for clarity of language within the site by requesting more examples and information on presented concepts. Participants also requested that more graphics and charts be included in the site. It could be presumed that this is for visual appeal, but it would also be that the increased level of appropriate charts and graphics will facilitate further explanation and clarity of concepts. Finally, while heuristic data revealed no problem areas within navigation, think-aloud response data indicated a modification to the menubar and "previous" and "next" buttons. Participants st ated that page numbers as well as modules should be included on the main menubar to more easily enable users to move from one point in the site to the next.

Overall, it is clear that with some modification, the current website for this study will be an appropriate supplementary source for an onsite course in human performance technology. However, issues with respect to user learning gains need to be addressed before the site can be utilized as a distance delivered course. Further, the evaluation methodology utilized within this project appears to be highly successful in indicating navigation, attitudinal, informational and motivational issues within instructional websites. It is however still not clear as to how accurately learning gains can be measured using the same methodology. Further investigation of the methodology altering the number of participants used to measure learning gains is recommended in the future.

References

Bellotti, V. (1988). Implications of current design practice for the use of HCI techniques. In D.M. Jones & R. Winder (Eds.), People and Computers IV. Cambridge: Cambridge University Press, pp. 13-34.

Cory, M.D., Frick, T.W., & Hansen, L. (1997). User-centered design and usability testing of a web site: An illustrative case study. Educational Technology Research and Development, 45(4), 65-76.

Cummings, O. (1998). What stakeholders want to know. In S.M. Brown & J. Seidner (Eds.), Evaluating Corporate Training: Models and Issues. Boston: Kluwer Academic Publishers pp. 41-62.

Dick, W. & Carey, L. (1996). The Systematic Design of Instruction (4th ed.). New York: Harper Collins College.

Dumas, J.S. Redish, J.C. (1993). A Practical Guide to Usability Testing. Norwood: Ablex Publishing Corporation.

Hix, D & Hartson, H.R. (1994). Developing User Interfaces: Ensuring Usability Through Product and Process. New York: John Wiley & Sons Incorporated.

Molich, R., & Nielsen, J. (1990). Improving a human-computer dialogue, Communications of the ACM, 33(3), 338-348.

Nielsen, J. (1989b). Usability engineering at a discount. In G. Salvendy & M.J. Smith (Eds.), Designing and Using Human-Computer Interfaces and Knowledge Based Systems. Amsterdam: Elsevier Science Publishers, pp. 394-401.

Nielsen, J. (1990a). Big paybacks from "discount" usability engineering, IEEE Software, 7(3), 107-108.

Nielsen, J. (1993). Usability Engineering. Boston: Academic Press.

Nielsen, J. (1994). Guerrilla HCI: Using discount usability engineering to penetrate the intimidation barrier. In R. G. Bias & D. J. Mayhew (Eds.), Cost-Justifying Usability. Boston: Academic Press.

Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen, and R.L. Mack (Eds.), Usability Inspection Methods. New York: John Wiley & Sons.

Nielsen, J. (1996). Top ten mistakes in web design [On-line]. Available: http://www.useit.com/alertbox/9605.html

Patterson, A. & Bloch, B. (1987). Formative evaluation: A process required in computer-assisted instruction. Educational Technology, 27(6), 26-30.

Rubin, J. (1994). Handbook of usability testing: How to plan, design, and conduct effective tests. New York: John Wiley & Sons, Incorporated.

Scriven, M. (1967). The methodology of evaluation (AERA Monograph series on curriculum evaluation, No. 1). Chicago: Rand McNally.

Smith, P.L. & Ragan, T.J. (2000). Instructional Design (2nd Ed). Upper Saddle River: Prentice-Hall Incorporated.

Smith, P. & Wedman, J.F. (1988). Read-Think-Aloud protocols: a new data-source for formative evaluation. Performance Improvement Quarterly, 1(2), 13-22.

Whiteside, J., Bennett, J., & Holtzblatt, K. (1988). Usability engineering: Our experience and evolution. In M. Helander (Ed.), Handbook of Human-Computer Interaction. North-Holland: Elsevier Science Publishers, pp. 791-817.